d. Given an image coordinate $(x_i,y_i)=(3,4)$, focal length f=5 cm, and object depth z=10 cm, calculate the real-world coordinates (x_c,y_c). (1.0)

e. Enlist the intrinsic and extrinsic camera parameters. (1.0)

f. For what purpose, following code in Python may work? (2.0)

```
import cv2
import numpy as np
import matplotlib.pyplot as plt
import torch
import torchvision.transforms as transforms
from torchvision import models
# Load a pre-trained deep learning model for monocular depth estimation
model = models.detection.maskrcnn_resnet50_fpn(pretrained=True)
model.eval()
# Function to convert 2D image to depth map (approximate 3D)
def estimate_depth(image_path):
# Load image
image = cv2.imread(image_path)
image_rgb = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
# Preprocess image for the model
transform = transforms.Compose([transforms.ToPILImage(),
transforms.Resize((256, 256)),
transforms.ToTensor()])
input_tensor = transform(image_rgb).unsqueeze(0)
# Forward pass through the model
with torch.no_grad():
depth_map = model(input_tensor)
# Convert tensor to numpy array
depth_map_np = depth_map[0]['masks'].squeeze().numpy()
# Normalize depth values
depth_map_np = (depth_map_np - np.min(depth_map_np)) /
(np.max(depth_map_np) - np.min(depth_map_np))
return depth_map_np, image_rgb
# Load image and estimate depth
image_path = "your_image.jpg"  # Replace with actual image path
depth_map, original_image = estimate_depth(image_path)
# Display results
fig, ax = plt.subplots(1, 2, figsize=(10, 5))
ax[0].imshow(original_image)
ax[0].set_title("Original Image")
ax[1].imshow(depth_map, cmap='jet')
ax[1].set_title("Estimated Depth Map")
plt.show()
```

***

b. Apply the four operations mentioned in 5 (a) above, on the given source image shown in Fig. 6 (a) with the given structuring element in Fig. 6 (b) and show the output image in four cases: (6.0)

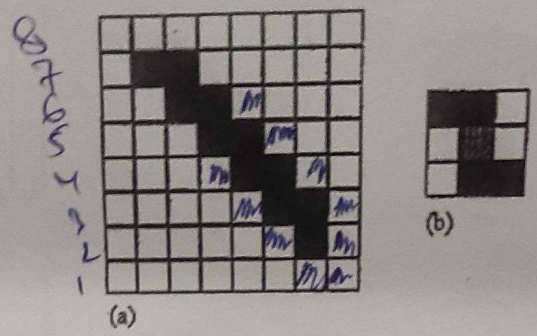

Fig. 6

c. What is meant by image warping? (1.0)

d. What is the difference between intra-class and inter-class variation of features. (1.0)

Q6.

a. With regard to Fig. 7, explain the following: (2.0)

i. Phantom points

ii. Stereo correspondence problem

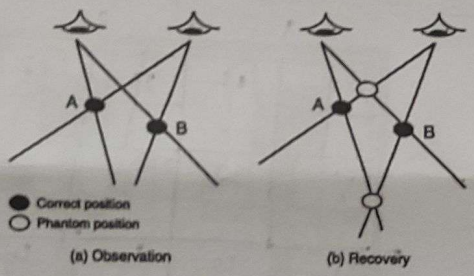

Fig.7

b. With regard to Figure 8, briefly explain Marr's vision paradigm: (2.0)

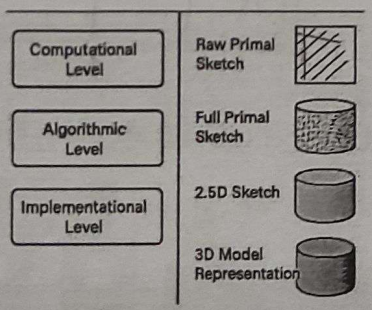

Fig. 8

c. If object A is located at 1.5 m from the observer and object B at 3 m, and both are viewed using the same stereo setup (baseline = 6 cm, focal length = 40 mm, pixel size = 0.004 mm), what are their disparities? Which object produces a larger disparity and what does that imply? (2.0)

e. Shown below in Fig.3 are the Ideal and Butterworth filters. (2.0)

i. Identify them as either Low Pass or High Pass.

ii. Compare their output response in terms of smoothing and noise removal.

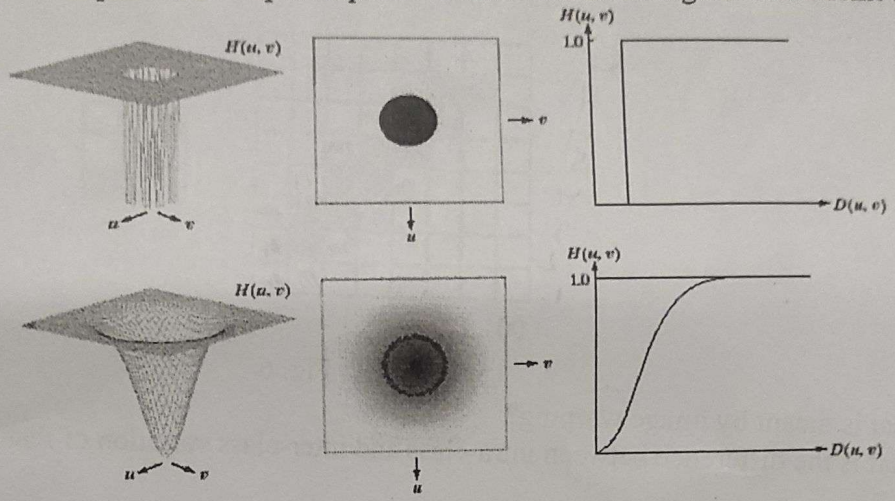

Fig. 3

Q4:

a. Name atleast two Large Vision Lanaguage Models used for segmentation. (1.0)

b. How does SAM work? Draw schematic. (2.0)

c. Write the QUADTREE Decomposition for the image patch (16x16) shown in Fig. 4. (5.0)

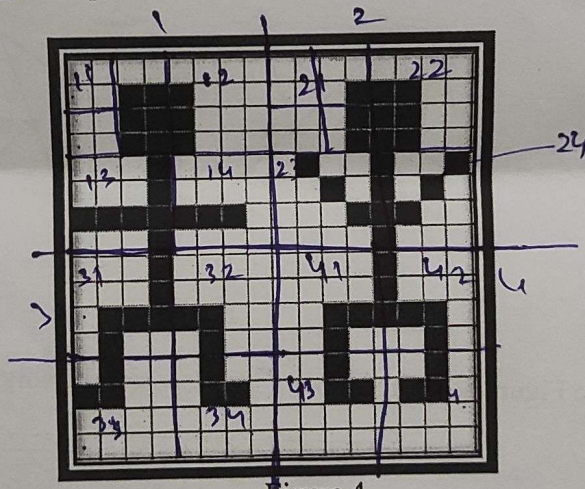

Figure 4

d. Calculate the EULER Number of image showin in Fig. 5 (2.0)

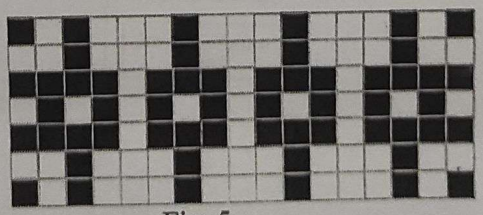

Fig. 5

Q5.

a. Write down the mathematical expressions for Erosion, Dilation, Opening, and Closing morphological operations. (2.0)

Q2:

a. Write the PROJECTION VECTORS for Fig.1. *You may assume BLACK: 0 and WHITE: 1* (2.0)

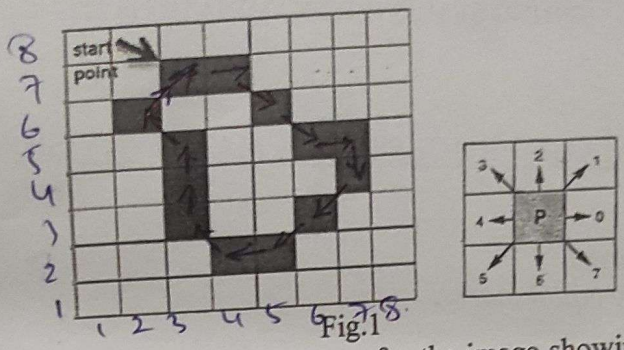

Fig.1

b. Write the RUN LENGTH ENCODING sequence for the image showin in Fig.1, mentioning the location and value of pixel at the location. (3.0)

c. Write the CHAIN CODE using 8-neighborhood for Fig.1 (2.0)

d. Calculate the approximate area of the image blob shown in Fig. 1 (including edge layer) with each pixel scaled as 1 inch by 1 inch. (*you don't need exact formula, use approximations*) (2.0)

e. Will the area increase or decrease if the pixel is located at centre of the each square? (1.0)

Q3.

a. Why does the CMYK model require a separate black (K) component despite theoretically being able to create black with cyan, magenta, and yellow? What practical issues does this resolve in printing? (2.0)

b. Explain how converting RGB to HSI can be advantageous in image processing tasks such as object detection and segmentation. In what scenarios would HSI be more useful than RGB? (2.0)

c. Briefly explain the HSI model shown in Fig.2 (2.0)

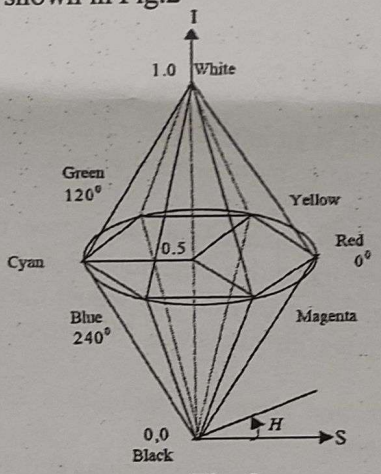

Fig. 2

d. An image of size 256 × 256 pixels is filtered in the frequency domain using a Gaussian lowpass filter with a cutoff frequency D_0=40. The Gaussian filter is defined by: (2.0)

$$H(u,v) = e^{-D^2(u,v)/2D_0^2}$$

where $D(u,v) = \sqrt{(u-128)^2 + (v-128)^2}$ (i.e., distances are centered)

i. Compute the value of the Gaussian lowpass filter at the point (128,160)

ii. Interpret the meaning of the resulting value obtained in (i) in terms of % attenuation.

Seat No: ME-235009

NED UNIVERSITY OF ENGINEERING & TECHNOLOGY

MASTER OF ENGINEERING (MECHANICAL)
SPRING SEMESTER EXAMINATIONS 2025

Dated:31-MAY-2025

Time: 3 Hours **Max.Marks:60**

DIGITAL IMAGE PROCESSING & MACHINE VISION - MC-541

Instructions:
- *Attempt all questions.*
- *All questions carry equal marks.*

Q1: Mark as True or False: (10.0)

i. Low-pass filters in digital image processing are primarily used to enhance fine details and sharpen images.
ii. In the spatial domain, convolution with a large kernel has a computational complexity of O (N^2 x M^2), while FFT-based filtering reduces this to O (N^2 x $\log_{10}$ N), making it dramatically faster for large-scale operations.
iii. The median filter is used for noise reduction, particularly for salt-and-pepper noise, and generally smooths edges rather than sharpening them.
iv. In machine vision, geometric transformations such as scaling, rotation, and translation only affect pixel locations but do not alter pixel intensity values.
v. Spatial resolution refers to the number of bits used to represent pixel intensity values in a digital image.
vi. Color accuracy depends on quantization (bit depth), not resolution.
vii. Replacing ambient lighting with carefully controlled lighting increases the complexity of subsequent subsystems.
viii. Object surface finish is an inherent scene constraints.
ix. Preprocessing is used to modify pixel values, altering the fundamental structural information present in the image.
x. Backlighting is used to enhance surface texture, making small features more distinguishable
xi. Spatial resolution refers only to the number of bits used per pixel.
xii. A grayscale image has pixel intensity values ranging from 0 to 255. If an image is thresholded at T = 150, all pixels with intensity ≤ 150 will be converted to white, while all pixels > 150 will be black.
xiii. Canny edge detection algorithm includes a smoothing step using a Gaussian filter before computing gradients.
xiv. A 5 × 5 kernel requires 25 multiplications and 24 additions per pixel during convolution.
xv. Histogram equalization does not necessarily brighten an image.
xvi. Otsu thresholding works by maximizing inter-class variance.
xvii. The sum of filter coefficients does not always need to be 1.
xviii. A Gaussian smoothing filter produces less blurring compared to an averaging filter and gives more importance to the central pixels.
xix. Transfer learning is ineffective for image denoising because pre-trained models are designed for classification tasks rather than noise removal
xx. The Structural Similarity Index (SSIM) is a better metric than Peak Signal-to-Noise Ratio (PSNR) for evaluating denoising performance because it is more aligned with human perception.